(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 363 037**

(51) Int. Cl.:
G10L 15/18 (2006.01)
G10L 15/22 (2006.01)
G10L 11/04 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **07380260 .5**
(96) Fecha de presentación : **21.09.2007**
(97) Número de publicación de la solicitud: **2040250**
(97) Fecha de publicación de la solicitud: **25.03.2009**

(54) Título: **Control de vehículos.**

(45) Fecha de publicación de la mención BOPI:
19.07.2011

(45) Fecha de la publicación del folleto de la patente:
19.07.2011

(73) Titular/es: **THE BOEING COMPANY**
100 North Riverside Plaza
Chicago, Illinois 60606-2016, US

(72) Inventor/es: **San-Segundo Hernandez, Ruben;**
Ferreiros Lopez, Javier;
Scarlatti, David;
Perez Villar, Victor y
Molina, Roberto

(74) Agente: **Ungría López, Javier**

DESCRIPCIÓN

Control de vehículos

Campo de la invención

La presente invención se refiere a control de vehículos activado por voz, y al control de UAV (vehículos aéreos no tripulados) utilizando voz en particular.

Antecedentes de la invención

Actualmente la tecnología de voz ha alcanzado un alto nivel de rendimiento y esto ha llevado a su mayor utilización en muchos sistemas críticos. Las investigaciones realizadas por compañías aeronaúticas e instituciones reguladoras en colaboración con grupos de expertos en tecnología de voz han presenciado el desarrollo de grandes bases de datos de voz y texto, junto con los nuevos modelos de procesamiento de voz y texto que se adaptan a los requisitos específicos del campo. Un área importante de aplicación crítica que puede beneficiarse de estas capacidades es el control de vehículos aéreos. En particular se benefician el control del tráfico aéreo (ATC) y las interfaces para UAV . Los UAV son de particular interés para la presente invención.

Cuando se desarrolla una interfaz de control UAV, es habitual incluir los siguientes módulos de reconocimiento de voz: un reconocedor de voz para convertir habla natural en una secuencia de palabras, un módulo de comprensión del lenguaje natural que extrae los conceptos semánticos principales del texto (los comandos que van a ejecutarse y sus datos correspondientes para control UAV), y un módulo de generación de respuestas para crear una respuesta natural para el piloto que se convertirá en voz mediante un sintetizador de voz. La respuesta confirma el comando recibido.

El software de reconocimiento de voz que se ha desarrollado hasta ahora se basa en dos fuentes de conocimiento, modelado acústico y modelado de lenguaje. En relación con el modelado acústico, los sistemas actuales de reconocimiento de voz se basan en modelos ocultos de Markov (HMM). Para cada alófono (una pronunciación característica de un fonema), se calcula un modelo HHM como resultado de un proceso de entrenamiento llevado a cabo utilizando una base de datos de voz. Una base de datos de voz consiste en varias horas de habla transcrita (compuesta de archivos con voz y texto combinado, en los que es posible correlacionar la señal de voz con las palabras pronunciadas por la persona). El tamaño de la base de datos determina la versatilidad y robustez del reconocimiento de voz. La adquisición de la base de datos es un proceso muy costoso porque requiere expertos en lingüística para transcribir a mano el habla pronunciada por diferentes hablantes.

El modelado de lenguaje complementa al modelado acústico con la información acerca de las secuencias de palabras más probables. Existen varias técnicas para el modelado de lenguaje incluyendo modelado de lenguaje basado en la gramática y modelado de lenguaje estadístico (N-gramas).

El modelado de lenguaje basado en la gramática consiste en definir todas las frases posibles que el sistema puede reconocer. Cualquier otra secuencia de palabras, no prevista en estas frases, se rechaza. Este modelo es más fácil de generar por una persona no experta, pero es muy estricto y no trata bien el habla espontánea o enfatizada que se produce en situaciones de la vida real.

El modelado de lenguaje estadístico consiste en calcular la probabilidad de una palabra, dadas las N-1 palabras anteriores. Por ejemplo, un modelo de 3-gramas consiste en las probabilidades de cada palabra posible precedida por cualquier combinación de dos palabras. El modelo estadístico se genera automáticamente a partir de algún texto orientado a la aplicación (conjunto de frases), considerando un proceso de suavizado para secuencias no vistas. Este suavizado significa que en cierta se permiten todas las secuencias de palabras medida (no hay secuencias de palabras prohibidas), satisfaciendo el papel de un factor de robustez fundamental. Este hecho es muy importante cuando se modela habla espontánea puesto que contiene repeticiones de palabras, dudas, etc.

Hasta ahora, todos los sistemas de reconocimiento de voz incorporados en interfaces UAV son programas comerciales tales como los proporcionados por Microsoft™ y Nuance™. El desarrollador de la interfaz UAV, normalmente un experto en la asignación de tareas UAV y en pilotaje pero que no es necesariamente un experto en tecnología de voz, integra estos reconocedores de voz. Aunque estos sistemas de reconocimiento de voz están evolucionando a motores de software más fáciles de utilizar y más robustos, todavía hay importantes limitaciones en su configuración que afectan drásticamente al rendimiento del reconocimiento de voz. Un aspecto importante es el modelado de lenguaje: los motores comerciales de reconocimiento ofrecen la posibilidad de definir un modelo basado en la gramática (fácil de definir por una persona no experta), pero esta configuración no es lo suficientemente flexible para habla espontánea o enfatizada que aparece a menudo en interfaces de control UAV.

Para entender comandos hablados, se debe extraer la información semántica o “significado” (dentro del dominio de

aplicación específica) de la salida del reconocedor de voz (es decir, la secuencia de palabras que proporciona). La información semántica puede representarse por medio de una trama que contiene algunos conceptos semánticos. Un concepto semántico consiste en un identificador o atributo y un valor. Por ejemplo, un concepto podría ser "CÓDIGO_PUNTO_DE_REFERENCIA" mientras que el valor es "A01". Normalmente, la compresión del lenguaje natural se realiza mediante técnicas basadas en reglas. Estas relaciones entre conceptos semánticos y secuencias de palabras y otros conceptos se definen a mano por un experto. Las técnicas basadas en reglas pueden clasificarse en dos tipos, estrategias descendentes (*top-down*) y ascendentes (*bottom-up*).

En una estrategia descendente, las reglas buscan conceptos semánticos a partir de un análisis global de la frase completa. Esta estrategia intenta hacer corresponder todas las palabras de la frase con una secuencia de conceptos semánticos. Esta técnica no es lo suficientemente flexible y robusta para tratar errores en la secuencia de palabras proporcionadas por el reconocedor de voz. Incluso un único error puede provocar que falle el análisis semántico. La mayoría de intentos anteriores en interfaces de voz para mando y control UAV utilizan técnicas basadas en reglas con estrategia descendente.

En una estrategia ascendente, el análisis semántico se realiza empezando por cada palabra de manera individual y extendiendo el análisis a palabras de contexto vecinas y otras islas conceptuales ya construidas. Esta extensión se realiza para encontrar combinaciones específicas de palabras y/o conceptos (bloques) que generan un concepto semántico de nivel superior. Las reglas implementadas por el experto definen estas relaciones. Esta estrategia es más robusta frente a los errores de reconocimiento de voz y es necesaria cuando se utiliza un modelo de lenguaje estadístico en el software de reconocimiento de voz.

El módulo de generación de respuestas traduce los conceptos entendidos a una frase del lenguaje natural utilizada para confirmar el comando de vuelta al piloto. Estas frases puede ser fijas o pueden construirse utilizando plantillas con algunos campos variables. Estos campos se rellenan con la información obtenida de la interpretación semántica de la frase anterior. Ambos tipos de módulos de generación de respuestas se han utilizado en el pasado para el mando y control UAV. Finalmente, la frase de lenguaje natural se convierte en voz por medio de un sistema de conversión de texto a voz que utiliza un sintetizador de voz.

Sumario de la invención

Contrariamente a estos antecedentes, y desde un primer aspecto, la presente invención reside en un procedimiento implementado por ordenador para controlar un vehículo que comprende: recibir una o más instrucciones emitidas como voz; analizar la voz utilizando software de reconocimiento de voz para proporcionar una secuencia de palabras y una medida de confianza de palabras para cada palabra así reconocida; analizar la secuencia de palabras para identificar un concepto semántico correspondiente a una instrucción basándose en el análisis y en un nivel de confianza semántica para el concepto semántico identificado, obtenido al menos en parte con referencia a las medidas de confianza de palabras de las palabras asociadas con el concepto semántico; proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica; y utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo.

La utilización de medidas de confianza es ventajosa al aumentar la precisión del reconocimiento de voz. Además, permite adaptar la confirmación hablada a la precisión anticipada del reconocimiento de voz, tal como se expresa a través de las medidas de confianza.

De manera opcional, analizar la voz para proporcionar una secuencia de palabras comprende utilizar modelos ocultos de Markov. Pueden utilizarse modelos ocultos de Markov continuos.

Preferiblemente, analizar la voz para proporcionar una secuencia de palabras comprende utilizar un modelo de lenguaje que proporciona información acerca de las secuencias más probables de palabras que esperan encontrarse. De manera opcional, el procedimiento puede comprender utilizar el modelo de lenguaje para proporcionar un modelo estadístico de 2-gramas. Pueden utilizarse otros modelos estadísticos de N-gramas.

De manera opcional, el procedimiento puede comprender analizar tanto la secuencia de palabras como los niveles de confianza de palabras asociados para identificar el uno o más conceptos semánticos.

Preferiblemente, el procedimiento comprende analizar la secuencia de palabras para identificar el uno o más conceptos semánticos utilizando un enfoque ascendente que empieza con un análisis de cada palabra identificada de manera individual y luego extendiendo el análisis a palabras vecinas. Esto es opuesto al enfoque descendente, menos preferido.

Opcionalmente, el procedimiento puede comprender analizar la secuencia de palabras para identificar el uno o más conceptos semánticos etiquetando cada palabra según su tipo, por ejemplo, comando, dígito o letra, y buscando patrones conocidos de la secuencia de etiquetas, permitiendo así la deducción de un concepto semántico. Por

ejemplo, una secuencia conocida de dígitos y letras (por ejemplo, letra-dígito-dígito) puede corresponderse con la convención de etiquetado de un comando particular (por ejemplo, los códigos de puntos de referencia se identifican de esta manera).

Por supuesto, habrá situaciones en las que el proceso de reconocimiento de voz tendrá dificultades, por ejemplo debido a ruido de fondo o instrucciones mal pronunciadas. Para ayudar en la robustez del sistema, puede utilizarse un umbral para probar un concepto semántico. La medida de confianza semántica puede compararse con el umbral y se realiza una acción dependiendo de si la medida supera o no el umbral. Por ejemplo, la confirmación hablada del concepto semántico identificado puede incluir una indicación de que la instrucción no se entendió cuando el nivel de confianza semántica está por debajo del umbral. El procedimiento puede comprender utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo sólo cuando el nivel de confianza semántica supera un umbral. El umbral puede fijarse de diferentes maneras. Puede corresponder a una constante, fijada por adelantado. Como alternativa, el umbral puede fijarse por un operador, y ajustarse tan frecuentemente como se desee. Como otra alternativa, el umbral puede ser una variable que se calcula como parte de un sistema de retroalimentación. Por ejemplo, el umbral puede variarse según cuántas de las confirmaciones habladas llevan a una corrección por parte del operador (muchas correcciones indicarían sobreconfianza y un umbral demasiado bajo).

De manera opcional el procedimiento puede comprender proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica y también un nivel de verbosidad indicado. Preferiblemente, la velocidad y/o tono de habla de la confirmación hablada aumenta a medida que disminuye el nivel de verbosidad indicado.

De manera similar, el procedimiento puede comprender proporcionar una confirmación hablada del concepto semántico identificado basándose en el nivel de confianza semántica y en un nivel de urgencia, disminuyendo la verbosidad de la confirmación hablada con un nivel de urgencia creciente. Preferiblemente, la velocidad y/o tono de habla de la confirmación hablada aumenta a medida que aumenta el nivel de urgencia.

El procedimiento puede comprender generar la confirmación hablada utilizando plantillas de respuesta. Las plantillas de respuesta pueden contener respuestas básicas a las que puede añadirse detalle: el concepto semántico identificado determinará la plantilla a utilizar y el detalle que ha de añadirse.

Desde un segundo aspecto, la presente invención reside en un aparato para controlar un vehículo, que comprende: una entrada dispuesta para recibir una o más instrucciones emitidas como voz; una memoria dispuesta para almacenar la voz recibida; un módulo de reconocimiento de voz dispuesto para analizar la voz para proporcionar una secuencia de palabras y una medida de confianza de palabras para cada palabra así reconocida; un módulo de comprensión del lenguaje natural dispuesto para recibir la secuencia de palabras y las medidas de confianza de palabras, y analizar la secuencia de palabras para identificar un concepto semántico correspondiente a una instrucción basándose en el análisis y en un nivel de confianza semántica para el concepto semántico identificado, obtenido al menos en parte con referencia a la medidas de confianza de palabras de las palabras asociadas con el concepto semántico; un módulo de generación de respuestas dispuesto para proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica; y un módulo comercial de generación dispuesto para utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo. La presente invención también reside en un programa informático para implementar los procedimientos anteriores y en un producto de programa informático que incluye un programa informático de este tipo.

Otras características preferidas, pero opcionales, se definen en las reivindicaciones adjuntas.

Breve descripción de los dibujos

Con el fin de que la presente invención pueda entenderse más fácilmente, a continuación se describirán realizaciones preferidas, sólo a modo de ejemplo, con referencia a los dibujos adjuntos, en los que:
la figura 1 es una representación esquemática de un procedimiento para controlar un vehículo utilizando reconocimiento de voz;
la figura 2 es una representación esquemática de un sistema para llevar a cabo el procedimiento mostrado en la figura 1;
la figura 3 es una representación esquemática de un procedimiento para entrenar el modelo de lenguaje de la figura 2;
la figura 4 es una representación esquemática de un procedimiento empleado por el módulo de comprensión de lenguaje natural de la figura 2;
la figura 5 es un boceto que muestra el control remoto de un UAV desde plataformas tanto con base en el aire como con base en tierra.

Descripción detallada de la invención

Un procedimiento general para controlar un vehículo, tal como un UAV, se muestra en la figura 1. En 10, se dice un comando que se recibe en 12 por un sistema dispuesto para implementar el procedimiento. El sistema puede ser, por ejemplo, un ordenador programado de manera adecuada. El sistema almacena la voz que se ha recibido en 12. Después, el sistema emplea algoritmos de reconocimiento de voz en 14 para identificar palabras a partir de la voz almacenada. Además, el sistema determina una medida de confianza de palabras para cada palabra identificada. En 16, el sistema identifica entonces conceptos semánticos a partir de las palabras identificadas previamente y las medidas de confianza de palabras en 16. El sistema también proporciona una medida de confianza semántica para cada concepto semántico identificado.

Después, el sistema genera una respuesta verbal en 18 y utiliza un sintetizador de voz en 20 para reproducir de manera audible la respuesta en forma hablada. La respuesta se corresponde con una confirmación del comando tal como lo entendió el sistema. Se reproduce de manera audible de manera que puede tomarse una acción correctiva si el comando se ha entendido mal, o como una petición de una entrada adicional cuando el sistema no ha podido entender el comando.

Además, el sistema utiliza los conceptos semánticos identificados para generar un código de comando en 22. Este código de comando puede retransmitirse al sistema de control del vehículo, efectuando de ese modo el control del vehículo.

La figura 2 muestra funcionalmente un sistema 24 para implementar el procedimiento descrito anteriormente. El sistema 24 se describe en el contexto del control remoto de un UAV utilizando un ordenador portátil. El ordenador portátil se programa de manera adecuada, como se describirá posteriormente. Se apreciará que este escenario no es más que una realización particular de la invención y que se prevén muchas otras aplicaciones.

El sistema comprende un número de módulos, dispuestos como sigue.

El módulo 26 de reconocimiento de voz convierte el habla natural recibida en una secuencia de palabras (texto). Una característica importante de este módulo 26 es el modelo de lenguaje estadístico que se ha entrenado para aumentar la robustez frente al habla espontánea. Otra característica importante es la estimación de confianza: cada palabra reconocida se etiqueta con un valor de confianza de palabras que representa la convicción del módulo 26 de reconocimiento de voz acerca de la validez de su propio trabajo. Se utilizan valores de confianza entre 0,0 (confianza más baja) y 1,0 (confianza más alta). Los valores de confianza de palabras se corresponden con la fiabilidad de la secuencia de palabras obtenida a partir del módulo 26 de reconocimiento de voz.

El módulo 28 de comprensión del lenguaje natural extrae los conceptos semánticos principales (comandos y sus correspondientes datos) del texto, utilizando reglas semánticas definidas por un experto. Este módulo 28 también genera una estimación de confianza para cada concepto semántico extraído.

El módulo 30 de generación de respuesta utiliza varias plantillas de respuesta para crear una frase de lenguaje natural como confirmación del comando entendido. Las plantillas de respuesta utilizadas son variables y dependen de los valores de confianza de los conceptos semánticos, del estado de urgencia y de la verbosidad de confirmación deseada. La frase se pasa a un módulo 62 sintetizador de voz, en el que se reproduce como voz.

La unidad 32 de generación de comandos utiliza el texto proporcionado por el módulo 28 de comprensión del lenguaje natural para obtener el código de comando. Este código de comando se pasa al UAV y se utiliza por el sistema de gestión de vuelo de UAV y por tanto determina aspectos del vuelo del UAV.

A continuación sigue una descripción más detallada del sistema de la figura 2.

El módulo 26 de reconocimiento de voz es un sistema de reconocimiento de voz del estado de la técnica que utiliza los HMM (Modelos ocultos de Markov) con las siguientes características principales.

Se utiliza un sistema de reconocimiento de voz continua que reconoce sonidos formados por varias palabras que se dicen de manera continua. En esta realización, el tamaño del vocabulario es de noventa y tres palabras.

El módulo 26 de reconocimiento de voz es independiente del hablante. Ha sido entrenado con una gran base de datos, haciéndolo robusto frente a una gran gama de hablantes potenciales sin entrenamiento adicional por parte de usuarios reales.

Además, el módulo 26 de reconocimiento de voz puede generar una secuencia de palabras óptima (dados los modelos acústicos y de lenguaje), una solución expresada como un gráfico acíclico dirigido de palabras que puede recopilar diferentes alternativas, o incluso las N mejores secuencias de palabras clasificadas por similitud con los

sonidos pronunciados.

El módulo 26 de reconocimiento de voz también proporciona una medida de confianza para cada palabra reconocida en la secuencia de palabras, con valores entre 0,0 y 1,0. Esta medida es importante porque el rendimiento del módulo 26 de reconocimiento de voz puede variar dependiendo de varios aspectos que incluyen el nivel de ruido del entorno, hablantes no nativos, habla más o menos espontánea, o la similitud acústica entre diferentes palabras contenidas en el vocabulario. Actualmente, los motores de reconocimiento de voz comerciales no proporcionan esta característica porque es difícil de gestionar cuando se diseña una interfaz de voz para aplicaciones críticas.

El módulo 26 de reconocimiento de voz se refiere a una base 34 de datos de modelos acústicos que almacena 3.500 HMM de tres fonemas para modelar todos los posibles alófonos y su contexto. La base 34 de datos de modelos acústicos también presenta dieciséis HMM de silencio y sonido para detectar efectos acústicos (casos no relacionados con el habla, como ruido de fondo, artefactos del hablante, pausas con marcadores de dudas,…) que aparecen en el habla espontánea. Es importante detectarlos y procesarlos con el fin de evitar estos ruidos que afectan al rendimiento del reconocimiento.

En esta realización, el módulo 26 de reconocimiento de voz utiliza HMM continuos. Esto significa que las funciones de densidad de probabilidad utilizadas en cada estado de cada modelo son funciones continuas (multigausianas). Este modelado ha mostrado ser la estrategia más potente para implementar los HMM. Los HMM acústicos se han entrenado con una base de datos muy grande, que contiene más de veinte horas de voz de 4.000 hablantes. El tamaño de la base 34 de datos y la variedad de los hablantes dota a los modelos 34 acústicos de una importante robustez y potencia de reconocimiento.

El módulo 26 de reconocimiento de voz también se refiere a una base 36 de datos que almacena un modelo de lenguaje. El modelo de lenguaje complementa a los modelos acústicos con información acerca de las secuencias de palabras más probables. En esta realización, el modelo 36 de lenguaje utiliza un modelado de lenguaje estadístico de 2-gramas. Este tipo de modelo calcula las probabilidades de que una palabra aparezca precedida de otra palabra. Como se observó anteriormente, este tipo de modelado de lenguaje presenta la mejor robustez cuando se modela habla espontánea (repeticiones de palabras, dudas,…), porque no excluye ninguna secuencia de palabras. Por otro lado, necesita una configuración más compleja de las herramientas automáticas para la generación del modelo de lenguaje, requiriendo la intervención de expertos.

La figura 3 muestra el procedimiento empleado para entrenar un modelo de lenguaje de 2-gramas basado en palabras a partir de la descripción de comandos originales proporcionada por los expertos de control UAV. Este procedimiento consiste en tres etapas.

El procedimiento comienza cuando se reciben comandos originales. Como primera etapa en la expansión de comandos, el módulo 38 de expansión de comandos expande cada descripción de comandos con referencia a una base 40 de datos de lista de comandos y considerando su estructura definida. Algunos ejemplos de expansión son los siguientes.

Para partes opcionales, el comando se expande considerando todas las estructuras posibles. Por ejemplo, ASCENDER [Y MANTENER] {$ALTURA} se expande como “ASCENDER {$ALTURA}” y “ASCENDER Y MANTENER {$ALTURA}” (las palabras entre llaves puede variar según lo que se haya especificado).

Para una lista de elementos, cuando se define una lista de posibles valores, se generan copias del mismo comando escogiendo un valor para cada instancia. Por ejemplo, (“RADIO CORTO | MEDIO | LARGO)”, proporciona tres ejemplos con valores diferentes generados (los elementos de la lista se expresan entre paréntesis y separados por líneas verticales).

Para una expansión de macros, cada macro se expande reproduciendo su estructura. Por ejemplo {$ALTURA} puede expandirse a varias estructuras:
“{$DÍGITO}{$ DÍGITO }{$DÍGITO}{$ DÍGITO } PIES” o “NIVEL DE VUELO {$DÍGITO}{$DÍGITO}{$DÍGITO}”, donde “{$DÍGITO}” es otra macro que contiene las palabras para los dígitos básicos desde “CERO” hasta “NUEVE”. Otro ejemplo pede ser {$CÓDIGO_PUNTO_DE_REFERENCIA} que podría expandirse a “{$LETRA}{$DÍGITO}{$DÍGITO}”.

Esta expansión de comandos presenta una importante limitación. Existen varios casos en los que no es posible expandir todos los valores posibles (letras o dígitos). Por ejemplo, si se quisiera expandir todos los posibles valores para un punto de referencia (considerando que estaría compuesto por latitud *dígito dígito* grados *dígito dígito* minutos *dígito dígito* segundos más longitud *dígito dígito dígito* grados *dígito dígito* minutos *dígito dígito* segundos) considerando todos los posibles valores de “dígitos”, existen 10x10 x 10x10 x 10x10 x 2x10x10 x 10x10 x 10x10 = $2X10^{11}$ posibilidades. Con el fin de evitar esta situación, se han considerado dos clases de palabras: letra y dígito, que entrenan un modelo de lenguaje (LM) basado en clases.

A continuación, en el procedimiento de la figura 3, el módulo 42 de entrenamiento de LM basado en clases genera un LM de 2 gramas de clases calculando las probabilidades de cualquier palabra/clase seguida por cualquier palabra/clase, considerando el comando parcialmente expandido en la etapa anterior. En este caso, hay dos clases de palabras: “letra” (con todas las posibles letras) y “dígito” (con todos los posibles dígitos). Durante este proceso, se aplica suavizado para proporcionar alguna probabilidad a las secuencias de palabras/clases que no se han visto en los comandos expandidos utilizando una base 44 de datos de lista de comandos expandidos. Este suavizado puede controlarse y ajustarse para esta tarea.

En la última etapa, el módulo 46 de expansión de clases convierte el LM de clases en un LM de palabras. Este proceso se lleva a cabo sustituyendo las probabilidades estimadas para cualquier clase (“dígito”, por ejemplo) por las probabilidades para las palabras que pertenecen a esta clase (“cero, uno, dos,…”). Las probabilidades de las palabras se calculan considerando las probabilidades de las clases (obtenidas en la etapa anterior) y el número total de palabras que pertenecen a esta clase, con referencia a una base 48 de datos de clases de palabras. Al final del procedimiento, el LM de palabras de 2 gramas basado en palabras se guarda en la base 36 de datos de modelos de lenguaje como el que puede utilizarse directamente por el módulo 26 de reconocimiento de voz.

El módulo 28 de comprensión del lenguaje natural es responsable de la extracción de la información semántica o “significado” (dentro del dominio de aplicación específica) a partir de la salida del módulo 26 de reconocimiento de voz (secuencia de palabras). La información semántica se transporta mediante una trama que contiene conceptos semánticos. Un concepto semántico consiste en un identificador y un valor. Por ejemplo, el concepto VELOCIDAD tiene “VELOCIDAD” como un identificador/atributo mientras que un posible valor es “71,96 m/s”. En esta realización, se han identificado treinta y tres conceptos principales; veintidós comandos y los datos correspondientes asociados con los mismos. Internamente, el módulo 28 de comprensión del lenguaje natural gestiona otros conceptos intermedios que llevan la información semántica cuando se desarrolla a partir de la entrada (exclusivamente compuesta por palabras) a través de representaciones intermedias con una mezcla de palabras y conceptos (conceptos tanto internos como principales).

El módulo 28 de comprensión del lenguaje natural se ha implementado utilizando una técnica basada en reglas que considera una estrategia ascendente. En este caso, las relaciones entre conceptos semánticos y secuencias de palabras y/o conceptos se definen a mano empleando a un experto. En una estrategia ascendente, el análisis semántico se realiza empezando desde cada palabra de manera individual y extendiendo el análisis a palabras de contexto vecinas o conceptos ya formados. Esta extensión se realiza para encontrar combinaciones específicas de palabras y/o conceptos que generan otro concepto. No todas las palabras contribuyen (o con otros términos, necesitan estar presentes) para la formación de la interpretación final. Las reglas implementadas por el experto definen esas relaciones y se clasifican en una base 50 de datos de reglas semánticas. Esta estrategia es más robusta frente a los errores de reconocimiento de voz y se prefiere frecuentemente cuando se utiliza un modelo de lenguaje estadístico en el reconocimiento de voz. Dependiendo del alcance de las relaciones de palabras definidas por las reglas, es posible conseguir diferentes compromisos entre fiabilidad del concepto extraído (superior con longitudes superiores) y robustez frente a errores de reconocimiento (superior con longitudes inferiores).

El proceso de comprensión se lleva a cabo en dos etapas, tal como se muestra en la figura 4. Primero, el módulo 52 de etiquetado mapea cada palabra a una de varias etiquetas sintáctico-pragmáticas utilizando una base 54 de datos de etiquetas. Por ejemplo: CERO, UNO, DOS,… se asignan a la etiqueta “DÍGITO” (y ALFA, BRAVO, CHARLIE,… se mapean a una etiqueta “ELEMENTO_ALFABÉTICO”). Un ejemplo de múltiples etiquetas son las palabras “PATRÓN DE VUELO”. Posteriormente se etiquetan con las marcas COMANDO13 (para establecer el patrón de vuelo predefinido) y COMANDO 14 (para establecer un patrón de vuelo específico), a través del proceso de comprensión y dependiendo de los datos detectados, sólo se selecciona una de estas etiquetas.

A continuación, el módulo 50 de comprensión trabaja aplicando diferentes reglas almacenadas en una base 58 de datos de reglas. Cuando se implementan por el módulo 56 de comprensión, estas reglas convierten las palabras etiquetadas en valores y conceptos semánticos por medio de palabras (o conceptos) de agrupación y definiendo conceptos de nombres. Con el fin de ilustrar el proceso, considérese un ejemplo para detectar CÓDIGO_MISIÓN, CÓDIGO_PATRÓN y CÓDIGO_PUNTO_DE_REFERENCIA. Estos tres conceptos tienen la misma estructura de valor letra-dígito-dígito. Hay una regla que detecta estos patrones en la secuencia de palabras y los sustituye por un CÓDIGO_GENERAL de concepto interno con un valor de código desarrollado a través de la concatenación de los bloques con las etiquetas especificadas. Este nuevo CÓDIGO_GENERAL se utiliza en este nivel cuando es necesario para más información para determinar completamente la naturaleza real de este valor. Por ejemplo, “bravo cero tres” se convierte más convenientemente en “B04”, y el CÓDIGO_GENERAL se renombra dependiendo del código detectado. “B04” puede corresponder a un código de misión, en cuyo caso CÓDIGO_GENERAL se convierte en CÓDIGO_MISIÓN.

Como se mencionó, el módulo 28 de comprensión del lenguaje natural genera un valor de confianza semántica para cada concepto identificado. Los valores varían entre 0,0 (confianza más baja) y 1,0 (confianza más alta). Este valor

de confianza se calcula mediante un procedimiento interno que se codifica dentro del intérprete de lenguaje propietario que ejecuta cada regla.
En este motor interno hay “funciones primitivas” responsables de la ejecución de las reglas escritas por los expertos. Cada primitiva tiene su propia manera de generar la confianza para los elementos que produce. Un caso común para las primitivas es que comprueben la existencia de una secuencia de bloques semánticos para generar algunos nuevos, en el que la primitiva asigna normalmente a los bloques recién creados la confianza media de los bloques en los que se ha basado.

Por ejemplo, la medida de confianza del concepto CÓDIGO_GENERAL descrito anteriormente es la media de los valores de confianza de palabras para “BRAVO”, “CERO” y “TRES”. Después, CÓDIGO_MISIÓN tendrá el mismo valor de confianza que el concepto CÓDIGO_GENERAL. En otros casos más complejos, la confianza para los nuevos bloques puede depender de una combinación de confianzas de una mezcla de palabras y/o conceptos internos o finales.

En aplicaciones críticas como interfaces de control UAV, es muy importante conseguir un alto nivel de rendimiento pero también es muy útil tener una medida de confianza que proporcione información acerca de la fiabilidad de la información semántica obtenida. Estas medidas evitan ejecutar acciones UAV con información posiblemente mal entendida, aumentando la fiabilidad de todo el sistema 24.

En esta realización, el módulo 30 de generación de respuestas utiliza plantillas de respuesta almacenadas en una base 60 de datos de tramas de respuesta para crear una frase del lenguaje natural como confirmación del comando entendido. En esta realización, las plantillas 60 de respuesta son variables y dependen de las confianzas de conceptos semánticos, de la urgencia UAV y de la verbosidad de la confirmación. En el módulo 30 de generación de respuestas, se han definido tres tipos de plantillas 60 de respuesta correspondientes a tres niveles de verbosidad.
Para “larga”, el sistema genera las frases más largas que incluyen toda la información entendida.

Para “corta”, la frase se acorta y parte de la información se omite. Las partes más largas y más tediosas se omiten, es decir, aquéllas para las que las interfaces de voz se justifican peor, como longitudes o latitudes completamente especificadas que podrían confirmarse mejor de una forma textual o gráfica.

Para “m corta”, el sistema sólo afirma la comprensión del comando, sin ninguna especificación acerca de lo que se ha entendido realmente.

El nivel real de verbosidad se modula a través de la especificación de dos parámetros, estado del sistema de urgencia y verbosidad de confirmación deseada. Un nivel superior de urgencia implica menos verbosidad por sí misma, mientras que un nivel superior de verbosidad de confirmación aumenta el tamaño de la respuesta. Existen tres niveles de estados de urgencia (alta, media y baja) y tres niveles de verbosidad de confirmación (alta, media y baja). La tabla a continuación muestra el mapeo entre sus correspondientes ajustes y la verbosidad del sistema global.

URGENCIA

VERBOSIDAD	**baja**	**media**	**alta**
baja	corta	m corta	m corta
media	larga	corta	m corta
alta	larga	larga	larga

Además del nivel de verbosidad, la acción realizada por el UAV y los contenidos reales de la respuesta dependerán de la confianza de comprensión obtenida para el sonido actual, según se compare con un umbral de confianza. Adicionalmente, cuando el nivel de comprensión es superior al umbral de confianza, el sistema proporciona diferentes salidas dependiendo de la estructura del comando (si el comando contiene los datos correctos para ejecutarlo o no). La siguiente tabla muestra ejemplos de contenidos de salida para las tras longitudes de respuesta dependiendo de la confianza de comprensión y la completitud de los datos de comando.

LONGITUD DE RESPUESTA		por encima del umbral	por debajo del umbral e incorrecto	Por debajo del umbral y correcto
	larga	Misión 01 de actuación	Comando C2, activar comando pero datos no entendidos	Lo siento, comando no entendido. Por favor, repita
	corta	Misión 01	Comando incompleto, no ejecutado	Lo siento, comando no entendido.
	m corta	OK	No entendido	No entendido

La frase de lenguaje natural proporcionada por el módulo 30 de generación de respuestas se convierte en voz por medio de un módulo 62 de síntesis de voz. Este módulo 32 utiliza un algoritmo de concatenación de unidades diáfonas, que puede modificar la velocidad de habla y el tono del hablante. La velocidad de habla y el tono del hablante se han ajustado para cada longitud de respuesta definida como sigue.

Para respuestas largas, los valores por defecto son 180 sílabas / minuto (velocidad de habla) y aproximadamente 130 Hz (tono del hablante).

Para respuestas cortas, la velocidad de habla se aumenta un 10% y el tono también se aumenta un 10% para generar una voz más rápida y más dinámica.

Para respuestas m cortas, la velocidad de habla se aumenta un 25 % y el tono se aumenta un 20% a partir de los valores por defecto.

Haciendo referencia ahora al control real del UAV, el módulo 32 de generación de comandos recibe los conceptos semánticos y medidas de confianza semántica desde el módulo 28 de comprensión de lenguaje natural. Siempre que las medidas de confianza superen un umbral, el módulo 32 de generación de comandos convierte los conceptos semánticos en código de comando. Este código de comando se corresponde con las instrucciones que van a proporcionarse al sistema de gestión de vuelo del UEV para provocar que vuele según se desee. El código de comando se proporciona a un transmisor 64 que transmite el código para su recepción por el UAV.

El UAV 66 puede controlarse de manera remota desde una variedad de plataformas diferentes. La figura 5 muestra dos ejemplos de este tipo. Se muestra un ordenador 68 portátil en una estación de tierra que transmite códigos de comando al UAV 60. El enlace de comunicación puede ser bidireccional, es decir, el UAV 66 puede transmitir mensajes de vuelta a la estación 68 de tierra, por ejemplo, para confirmar la recepción de comandos. La figura 5 muestra una segunda plataforma 70 con base en el aire. En ella, un oficial de armas sobre un avión 70 utiliza un ordenador programado de manera adecuada para controlar el UAV 60. De nuevo, este enlace de comunicación puede ser bidireccional, como para el enlace de la estación de tierra. Por tanto, las dos aeronaves pueden realizar misiones juntas, por ejemplo realizando el UAV 66 un papel de reconocimiento para ubicar objetivos que luego se captan por el avión 70 tripulado. Por supuesto, tanto la plataforma 70 con base en el aire como la 68 con base en tierra pueden utilizarse juntas.

El experto en la técnica apreciará que pueden hacerse variaciones a la realización descrita anteriormente sin apartarse del alcance de la invención definida por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento implementado por ordenador para controlar un vehículo, que comprende:
recibir una o más instrucciones emitidas como voz;
analizar la voz utilizando software de reconocimiento de voz para proporcionar una secuencia de palabras y una medida de confianza de palabras para cada palabra así reconocida;
analizar la secuencia de palabras para identificar un concepto semántico correspondiente a una instrucción basándose en el análisis y en un nivel de confianza semántica para el concepto semántico identificado, obtenido al menos en parte con referencia a las medidas de confianza de palabras de las palabras asociadas con el concepto semántico;
proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica; y
utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo.

2. El procedimiento según la reivindicación 1, en el que analizar la voz para proporcionar una secuencia de palabras comprende utilizar modelos ocultos de Markov (continuos).

3. El procedimiento según la reivindicación 1 o la reivindicación 2, en el que analizar la voz para proporcionar una secuencia de palabras comprende utilizar un modelo de lenguaje que proporciona información acerca de las secuencias de palabras más probables que esperan encontrarse.

4. El procedimiento según la reivindicación 3, que comprende utilizar el modelo de lenguaje para proporcionar un modelo estadístico de 2-gramas.

5. El procedimiento según cualquier reivindicación anterior, que comprende analizar la secuencia de palabras y los niveles de confianza de palabras asociados para identificar el uno o más conceptos semánticos.

6. El procedimiento según cualquier reivindicación anterior, en el que el concepto semántico comprende un identificador y un valor.

7. El procedimiento según cualquier reivindicación anterior, que comprende analizar la secuencia de palabras para identificar el uno o más conceptos semánticos utilizando un enfoque ascendente que empieza con un análisis de cada palabra identificada de manera individual y luego extendiendo el análisis a palabras vecinas.

8. El procedimiento según cualquier reivindicación anterior, que comprende analizar la secuencia de palabras para identificar el uno o más conceptos semánticos etiquetando cada palabra según su tipo, por ejemplo comando, dígito o letra, y buscando patrones conocidos a partir de la secuencia de etiquetas, permitiendo así la deducción de un concepto semántico.

9. El procedimiento según cualquier reivindicación anterior, en el que proporcionar una confirmación hablada del concepto semántico así identificado incluye una indicación de que la instrucción no se entendió cuando el nivel de confianza semántica está por debajo de un umbral.

10. El procedimiento según cualquier reivindicación anterior, que comprende proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica y un nivel de verbosidad indicado.

11. El procedimiento según la reivindicación 10, que comprende proporcionar una confirmación hablada del concepto semántico identificado con una velocidad y/o tono de habla que se aumenta a medida que disminuye el nivel de verbosidad indicado.

12. El procedimiento según cualquier reivindicación anterior, que comprende proporcionar una confirmación hablada del concepto semántico identificado basándose en el nivel de confianza semántica y en un nivel de urgencia, disminuyendo la verbosidad de la confirmación hablada con el aumento del nivel de urgencia.

13. El procedimiento según la reivindicación 12, que comprende proporcionar una confirmación hablada del concepto semántico identificado con una velocidad y/o tono de habla que se aumenta a medida que aumenta el nivel de urgencia.

14. El procedimiento según cualquier reivindicación anterior, que comprende utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo sólo cuando el nivel de confianza semántica supera un umbral.

15. Un aparato para controlar un vehículo, que comprende:

una entrada dispuesta para recibir una o más instrucciones emitidas como habla;
una memoria dispuesta para almacenar el habla recibida;
un módulo de reconocimiento de voz dispuesto para analizar la voz para proporcionar una secuencia de palabras y una medida de confianza de palabras para cada palabra así reconocida;
un módulo de comprensión del lenguaje natural dispuesto para recibir la secuencia de palabras y las medidas de confianza de palabras, y analizar la secuencia de palabras para identificar un concepto semántico correspondiente a una instrucción basándose en el análisis y en un nivel de confianza semántica para el concepto semántico identificado, obtenido al menos en parte con referencia a las medidas de confianza de palabras de las palabras asociadas con el concepto semántico;
un módulo de generación de respuestas dispuesto para proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica; y
un módulo de generación de comandos dispuesto para utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo.

16. El aparato según la reivindicación 15, que comprende además una base de datos de modelos acústicos acoplados de manera operable al módulo de reconocimiento de voz, y en el que la base de datos de modelos acústicos presenta modelos ocultos de Markov almacenados en la misma, opcionalmente modelos ocultos de Markov continuos almacenados en la misma.

17. El aparato según la reivindicación 15 o la reivindicación 16, que comprende además una base de datos de modelos de lenguaje acoplada de manera operable al módulo de reconocimiento de voz, en el que la base de datos de modelos de lenguaje presenta datos almacenados en la misma que proporcionan información acerca de las secuencias de palabras más probables que esperan encontrarse.

18. El aparato según la reivindicación 17, en el que la base de datos de modelos de lenguaje presenta un modelo estadístico de 2-gramas almacenado en la misma.

19. El aparato según cualquiera de las reivindicaciones 15 a 18, en el que el módulo de comprensión del lenguaje natural está dispuesto para analizar la secuencia de palabras y los niveles de confianza de palabras asociados para identificar el uno o más conceptos semánticos.

20. El aparato según cualquiera de las reivindicaciones 15 a 19, en el que el módulo de comprensión del lenguaje natural está dispuesto para proporcionar el concepto semántico como un identificador y un valor.

21. El aparato según cualquiera de las reivindicaciones 15 a 20, en el que el módulo de comprensión del lenguaje natural está dispuesto para analizar la secuencia de palabras para identificar el uno o más conceptos semánticos utilizando un enfoque ascendente que empieza con un análisis de cada palabra identificada de manera individual y luego extendiendo el análisis a palabras vecinas.

22. El aparato según cualquiera de las reivindicaciones 15 a 21, en el que el módulo de comprensión del lenguaje natural está dispuesto para analizar la secuencia de palabras para identificar el uno o más conceptos semánticos etiquetando cada palabra según su tipo, por ejemplo, comando, dígito o letra, y para buscar patrones conocidos de la secuencia de etiquetas, permitiendo así la deducción de un concepto semántico.

23. El aparato según cualquiera de las reivindicaciones 15 a 22, en el que el módulo de generación de respuestas está dispuesto para proporcionar una confirmación hablada del concepto semántico así identificado que incluye una indicación de que la instrucción no se entendió cuando el nivel de confianza semántica está por debajo de un umbral.

24. El aparato según cualquiera de las reivindicaciones 15 a 23, en el que el módulo de generación de respuestas está dispuesto para proporcionar una confirmación hablada del concepto semántico así identificado basándose en el nivel de confianza semántica y un nivel de verbosidad indicado.

25. El aparato según la reivindicación 24, en el que el módulo de generación de respuestas está dispuesto para proporcionar una confirmación hablada del concepto semántico identificado con una velocidad y/o tono de habla que se aumenta a medida que disminuye el nivel de verbosidad indicado.

26. El aparato según cualquiera de las reivindicaciones 15 a 25, en el que el módulo de generación de respuestas está dispuesto para proporcionar una confirmación hablada del concepto semántico identificado basándose en el nivel de confianza semántica y en un nivel de urgencia, disminuyendo la verbosidad de la confirmación hablada con el aumento del nivel de urgencia.

27. El aparato según la reivindicación 26, en el que el módulo de generación de respuestas está dispuesto para proporcionar una confirmación hablada del concepto semántico identificado con una velocidad y/o tono de habla que

se aumenta a medida que aumenta el nivel de urgencia.

28. El aparato según cualquiera de las reivindicaciones 15 a 27, en el que la unidad de generación de respuestas presenta asociada con la misma una base de datos que contiene plantillas de respuesta para utilizar al construir la confirmación hablada.

29. El aparato según cualquiera de las reivindicaciones 15 a 18, en el que el módulo de generación de comandos está dispuesto para utilizar el concepto semántico así identificado para proporcionar una entrada de control para el vehículo sólo cuando el nivel de confianza semántica supera un umbral.

30. Un programa informático que comprende instrucciones de programa que, cuando se ejecutan, provocan que un ordenador opere según el procedimiento de cualquiera de las reivindicaciones 1 a 14.

31. Un producto de programa informático que contiene el programa informático de la reivindicación 30.

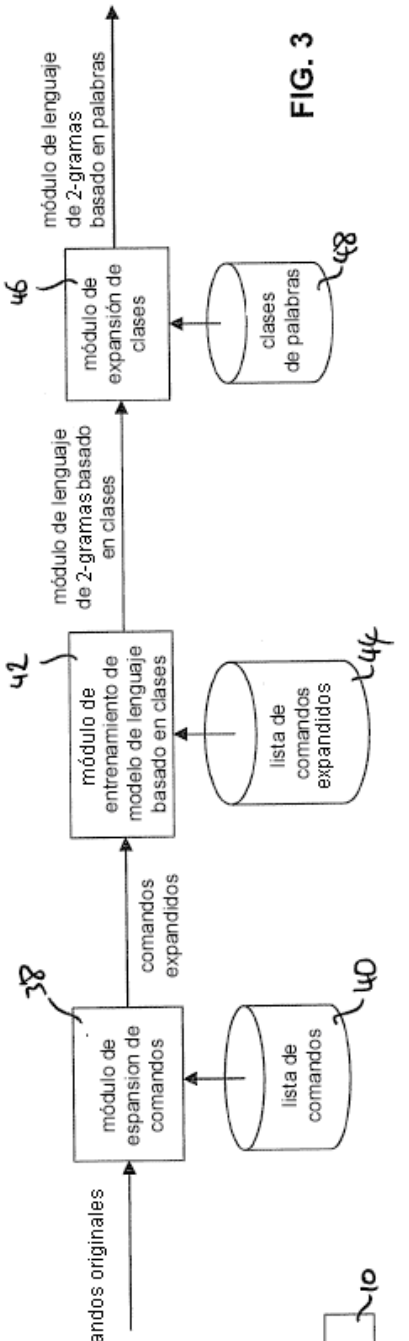

FIG. 3

FIG. 4

FIG. 1

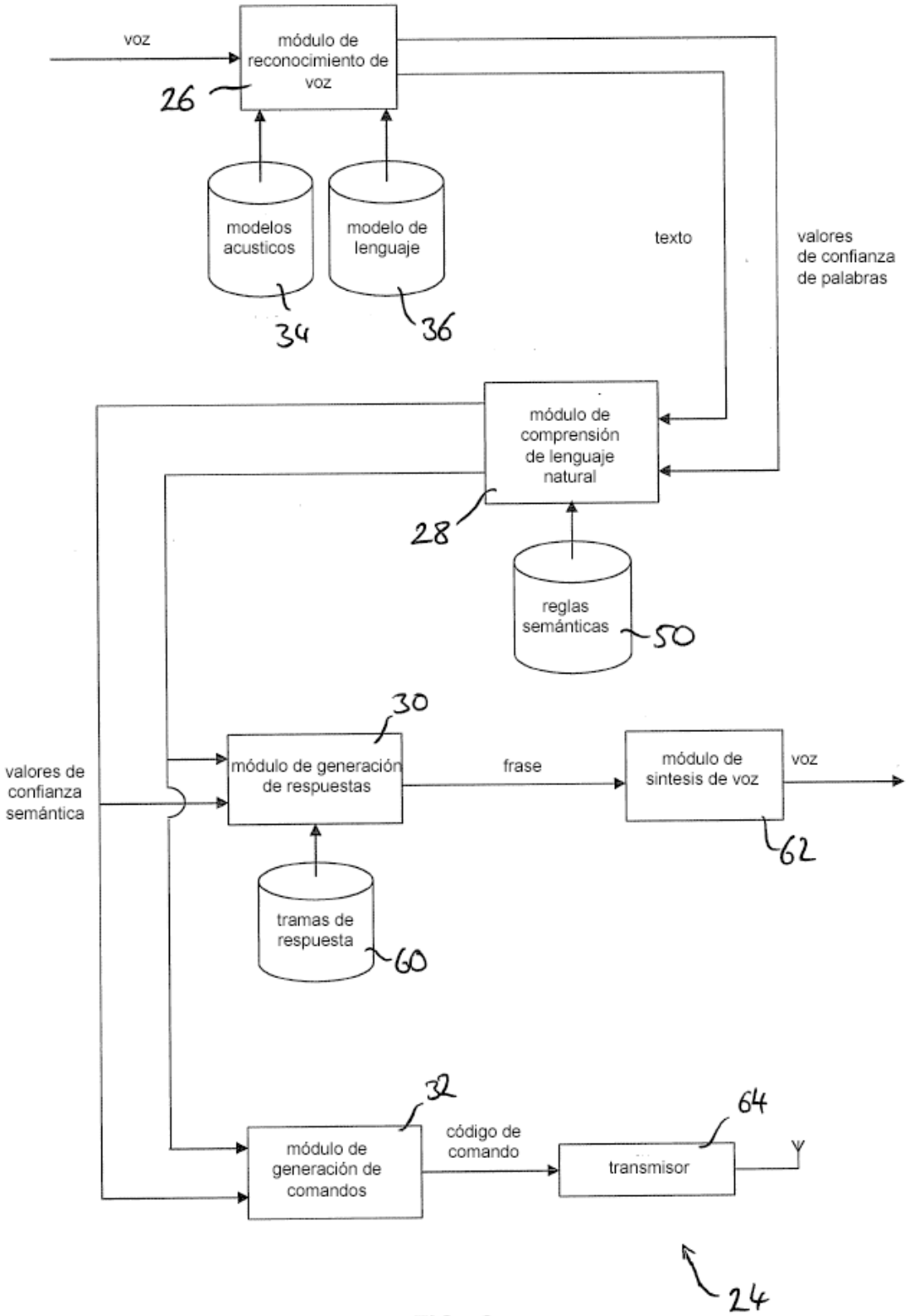
voz
módulo de reconocimiento de voz
26
modelos acusticos
modelo de lenguaje
34
36
texto
valores de confianza de palabras
módulo de comprensión de lenguaje natural
28
reglas semánticas
50
30
valores de confianza semántica
módulo de generación de respuestas
frase
módulo de sintesis de voz
voz
62
tramas de respuesta
60
32
módulo de generación de comandos
código de comando
64
transmisor
24

FIG. 2

Fig 5